

Amendments to the Specification:

Please replace paragraph [0054] on page 31 with the following replacement paragraph:

[0054] Figure 4 shows an example breakdown of this response time. The total response time $\ddot{A}T$ is made up of three different components:

- the time which server 14 uses to fulfill the client 10's request, i.e., web server processing time $\ddot{A}T_s$,
- the amount of time required for the request and associated response to be transmitted over network 12, i.e., the network transport time $\ddot{A}T_n$,
- the amount of time for client 10 to receive and process the response, i.e., the web client processing time $\ddot{A}T_c$.

Please replace paragraph [0055] on page 31 with the following replacement paragraph:

[0055] When a response time monitor is co-located with a server 14, it is relatively easy to monitor the web server processing time $\ddot{A}T_s$. For example, one can log the time when incoming requests are received by the server 14's firewall, and the time when the associated response to the request leave the firewall.

Please replace paragraph [0056] on pages 31-32 with the following replacement paragraph:

[0056] Network monitor 16 can get the network transport time (T_n), web processing time, web client processing time (T_c), retransmit time, dropped packets, and

network latency when it is co-located with the server 14 or when it is remotely located. Generally, the same techniques can be used in both the remote location context and in the locally connected context. When remotely monitoring delays associated with the response to a request, it is relatively easy to determine the total delay ΔT as well as the total delay attributable to the combination of web server processing time ΔT_s and network transport time ΔT_n ($\Delta T = \Delta T_s + \Delta T_n$). For example, a remote monitoring device can send a request to server 14 and receive a response, or it can monitor a request sent by another client 10 and log both the time at which the client sends the request and the time at which the client receives the response. [[ΔT_s +]] [[ΔT_n]] But using these techniques, there is no way for the remote monitor to determine which part of the total latency is attributable to web server response time ΔT_s and which part is attributable to network transport time ΔT_n . The present invention solves this problem by carefully monitoring different aspects of various transactions and deriving a measure of round-trip network latency. This round-trip latency parameter can be subtracted from the total latency to derive a measure of web server processing time ΔT_s = $\Delta T - \Delta T_n$.

Please replace paragraph [0057] on page 32 with the following replacement paragraph:

[0057] In more detail, a transaction-based protocol such as HTTP for example, uses a network transport protocol (e.g. TCP) to establish a session (e.g., a connection) between a web client and the web server. The time spent for this connection to complete may be termed TCP connect time (see Figure 5, block 50). This time calculation generally includes network latency and other possible network-induced delays that could occur. The remaining time, once these other

factors are discounted, is the TCP connect processing time (Figure 5, block 54)

(e.g., the time for establishing a TCP connection).